

AD 344487

OBSERVER PERFORMANCE WITH TV IMAGERY: GRAY SCALE AND RESOLUTION

D O C
RECEIVED
JUL 7 1972
R E G I S T E R E D
C

Reproduced by
NATIONAL TECHNICAL
INFORMATION SERVICE
U S Department of Commerce
Springfield VA 22151

DISTRIBUTION STATEMENT A
Approved for public release;
Distribution Unlimited

RECONNAISSANCE LABORATORY

MCDONNELL DOUGLAS

CORPORATION

**OBSERVER PERFORMANCE
WITH TV IMAGERY:
GRAY SCALE AND RESOLUTION**

by

S.H. Levine
R.A. Jauer
and
D.R. Kozlowski

MCDONNELL AIRCRAFT COMPANY

ABSTRACT

This study was performed as part of a McDonnell Aircraft Company funded research program to determine engineering design criteria for reconnaissance and air to air TV displays. The purpose of this experiment was to determine observer capability for identification of targets imaged on a simulated cockpit TV display. Observer performance was determined as a function of two display characteristics, shades of gray at the display face and TV lines over target. Six trained observers were presented TV images of 4 fighter aircraft at 3, 5, and 7 shades of gray, and 4, 8, 12, 16, and 20 TV lines over target. Fighter aircraft were presented one at a time against a uniform background. Performance was measured in terms of number correct identifications and speed of response. The correct response data showed reliable improvement in performance as gray shades were increased from 3 to 5 and 7 shades and no difference between 5 and 7 shades. Increases in TV lines over target yielded an increase in performance for both time and correct responses over the 4 to 12 lines range and no further performance increase from 12 to 20 TV lines.

Application of these data as engineering design criteria would indicate, within the limits of this study, that a TV cockpit display maintaining 5 shades of gray on the display face will yield a high level of observer performance. Increasing the shades of gray beyond 5 will yield little or no additional improvement in performance. A sensor-display system having 12 or more lines over target at the required identification range will be required to achieve high identification accuracy. Additional experiments to provide other design criteria for the definition of advanced TV systems are described.

TABLE OF CONTENTS

<u>Section</u>		<u>Page</u>
1	Introduction	1
2	Background	3
	Shades of Gray	3
	Number of TV Lines Over Target	5
3	Experimental Method	7
	Subjects	7
	Apparatus	7
	Procedure	11
4	Results	13
5	Discussion	23

LIST OF FIGURES

<u>Figures</u>		<u>Page</u>
1	Target Aspect Used in the Experiment	8
2	Block Diagram - Imaging Equipment	9
3	Percent and Number Correct for 3, 5 and 7 Shades of Gray	14
4	Percent and Number Correct Responses for TV Lines Over Target	15
5	Percent and Number Correct Responses for TV Lines Over Target by Gray Shades Interaction	17
6	Mean Time per Response - Gray Shades	18
7	Mean Time per Response - Lines Over Target	19
8	Mean Time per Response for TV Lines Over Target by Gray Shades Interaction	22

LIST OF TABLES

<u>Table</u>		<u>Page</u>
1	Analysis of Variance (Number of Correct Responses).	16
2	Analysis of Variance (Response Time).	20

LIST OF PAGES

Title Page
i
ii
iii
1 through 26

Section 1
INTRODUCTION

With the increase in use of airborne TV displays, engineering requirements for display design information are becoming more and more pronounced. This need is especially pressing in the area of cockpit displays because of the environmental restrictions placed on them and their proposed use for long range detection and identification of small complex targets. Among the environmental constraints inherent in cockpit displays is a limited brightness contrast range (gray scale rendition) on the display face, due to high ambient illumination in the open cockpit. Design criteria are needed to define the contrast range required on the display to maintain a high level of performance. Data on the number of TV lines over target* needed to initially achieve a high level of identification performance for complex targets will also be required to fully evaluate the effects of changes in the display brightness range. The present experiment, one of a series designed to generate human factors data for TV displays, is concerned with the effects of shades of gray and TV lines over target on target identification using a simulated cockpit TV display. Three, 5, and 7 shades of gray were used as the gray shade values most representative of the contrast range obtainable in a real display of this type. TV lines were varied from 4 to 20 in 4 line steps. This range of values encompasses a performance envelope ranging from chance to better than 80% correct performance.

* A TV line refers to a single raster line or single image forming horizontal sweep of the scanning beam across the display face.

REPORT H398
30 SEPTEMBER 1969

The purpose of this experiment is to determine the number of gray shades and TV lines over target required to insure adequate target identification performance on a cockpit type TV display.

Section 2

BACKGROUND

SHADES OF GRAY

Shades of gray is the measure of display brightness contrast used by the TV industry, and is equivalent to a $\sqrt{2}$ increase in brightness per gray shade. Gray scale is related to contrast by $N = 1 + 6.64 \log_{10} C$ where: N is the number of gray shades (usually rounded off to the highest integer value); C is the measured contrast ratio of the brightest to darkest area of the display image $\left(\frac{B_{max}}{B_{min}}\right)$. It should be noted that ambient illumination reflected from the display face is included in the measured contrast and contributes to Bmax and limits Bmin.

It is very difficult to present more than 5 to 7 shades of gray in an open cockpit TV display because of high daylight ambient illumination. It is important to note that all brightness differences are still present on the display, but the contrast ratio between the brightest and darkest is reduced. If the observer can discriminate between the reduced contrast differences, then the reduction in the number of gray shades should have little or no effect on his performance. Since gray shades are a physical measure of contrast range and not a measure of discriminable grays, it is possible that even a few shades of gray would be adequate for target identification, if other display quality parameters, such as signal-to-noise ratio, are adequate.

High ambient illumination and the concomitant reduction in shades of gray has been shown to greatly reduce target identification in cockpit displays. Erickson et. al. (1968)¹ demonstrated a severe decrement in performance for symbol identification when going from a darker cockpit to full sunlight. This,

however, represents the performance at the extremes of shades of gray values; as the conditions used by Erickson represent a low of less than 2 shades of gray and a high of more than 9 shades of gray.

Johns (1968)² imaging a terrain model on a cockpit display at 5, 7 and 9 shades of gray found marginal significance between 5 and 9 shades of gray in one phase of a two phase study. The second phase found no reliable difference between gray shades. The results of these two studies indicate that the decrement in performance found between low gray shade and high gray shade displays may not be distributed equally over the gray shade range. The decrement appears to occur at the low end of the range, probably at gray shades of 5 or less.

The problem of ambient illumination washing out gray shades rendition can be studied either by choosing a display and varying the ambient illumination or by manipulating the display to produce outputs analogous to the outputs which would be obtained using a variety of display types and ambient illuminations. The former method offers a high degree of face validity, but is limited to the specific display and illuminations used in the study. The latter method takes advantage of the fact that shades of gray are a product of ambient illumination and display capability, and further it treats the product of these two variables as an independent display variable. In determining engineering data for use in display specification we felt the best approach was one which allowed for the broadest scope of generalization. By choosing to treat the shades of gray directly as a variable, we are able to specify the required output independent of specific equipment or environmental restrictions.

NUMBER OF TV LINES OVER TARGET

The number of TV lines over target required for target identification is another important variable in the design of airborne TV displays. A display-sensor system must display the target with enough line coverage at the required range to ensure rapid identification of the target. The number of TV lines over target is a function of display size, lines in the full raster, and image scale. The TV display used was fixed with respect to display size and lines in the raster so that manipulating lines over target simulated the effect of scale change, independent of a sensor system. The actual air-to-air distance for any number of TV lines can be calculated from the scaling characteristics of the sensor system. Reducing the number of lines over target will also increase the difficulty of the identification task. The number of lines over target is thus an important display variable, both with respect to defining the air-to-air distance at which a target is identified and to determining the effect of varying identification difficulty on performance at the three gray shade levels.

Current literature^{1,3,4,5} indicates that identification performance will vary from chance to 80% or better using a range of values from 4 to 20 lines per target. This value^{1,3,5} is based on performance data acquired from a series of untrained observers detecting, recognizing and identifying, alphanumerics, symbols and simple silhouettes. The level of performance, for the most part, increases rapidly as lines increase from 0 to 8 lines over target, up to a performance level of about 80-90% correct. Performance increase then levels off with further increases in lines over target. The Elite Look Program,⁴ using simulated military targets as stimulus material, found performance at about 80%

REPORT H398
30 SEPTEMBER 1969

and still increasing slowly, with up to 20 lines over target. Of the studies concerned with the lines over target variable this one best simulated the actual target material which would be found in a real world display.

Section 3
EXPERIMENTAL METHOD

SUBJECTS

Six members of the McDonnell Aircraft Company Reconnaissance Laboratory were used as test subjects. All were either trained in image interpretation or experienced with real-time TV type raster displays.

APPARATUS

Twenty 35mm black and white slides showing four aircraft at five scales were made by suspending 1:72 scale models of a Mirage, AF-100 and two Mig 21s (FISHBED D and FISHBED E) with monofilament guys and photographing them against a kraft paper background. The imagery simulated an attitude of 40° right yaw, 9° roll and 10° downpitch (Figure 1). Positive transparencies at 1/4X, 1/2X, 3/4X, 1X and 1-1/4X were then mounted as slides and projected on a blue poly-coat rear projection screen. A TV camera viewed the projected image and transmitted it through a closed circuit system to a 9" TV monitor (Figure 2). TV camera and rear projection screen distance was set so the 1/4X slide image corresponded with 4 TV lines-over-target. The other four slide scales produced the required number of TV lines-over-target without any further change in the equipment. The visual angle subtense of the target was confounded with the increases in lines-over-target in much the same way it would be in an actual display. As lines-over-target increase, both target size and visual angle increase.



Figure 1 Target Aspect Used in the Experiment

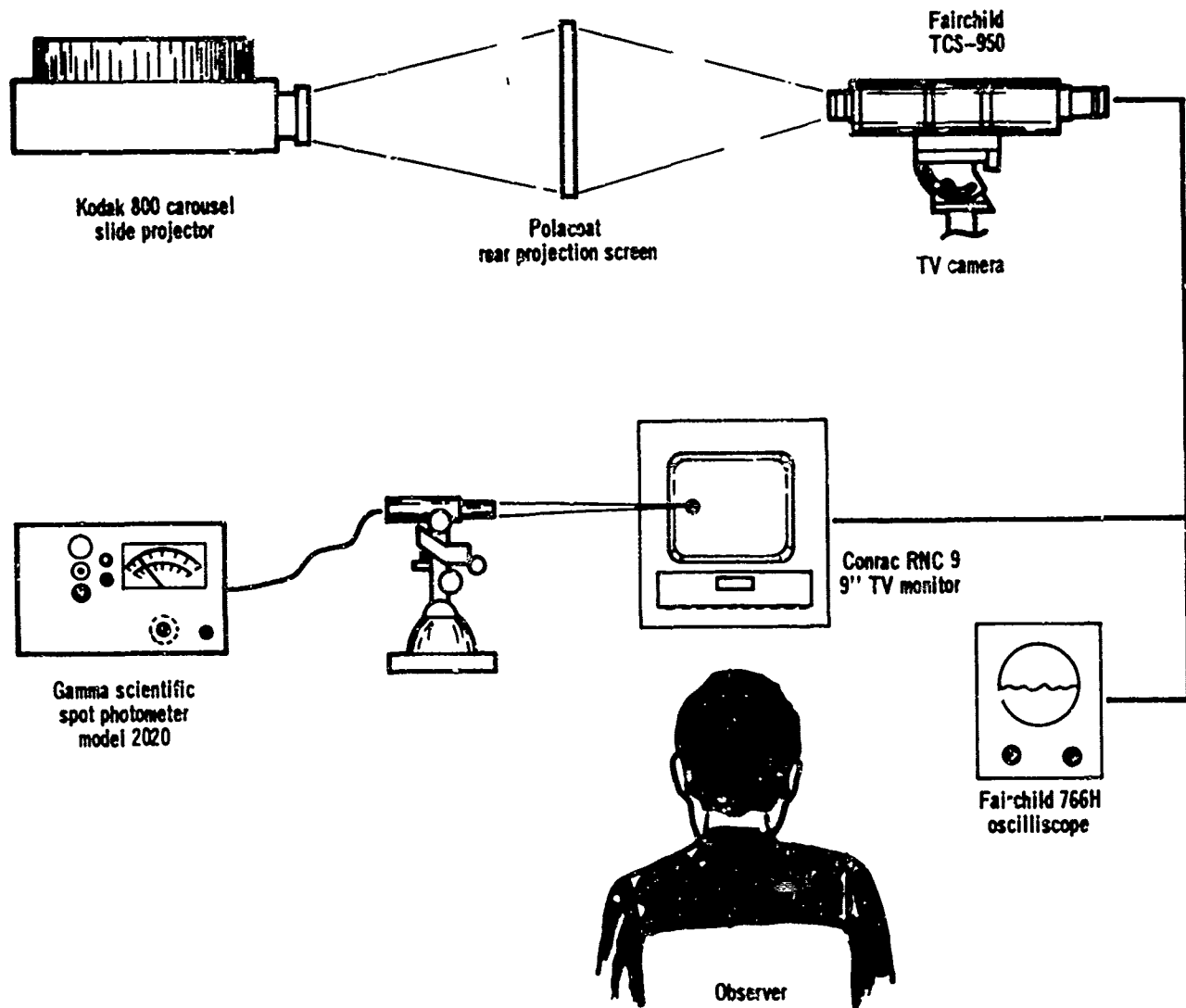


Figure 2 Block Diagram - Imaging Equipment Setup

An oscilloscope was used to adjust the video signal input to the display for maximum video and proper pedestal and sync levels that would permit continued proper synchronization. The display contrast or dynamic range was set by adjusting the brightness and contrast controls of the display to show 3, 5 or 7 gray shades. Each gray shade represents a $\sqrt{2}$ factor change in light intensity so that:

$$\frac{B_{\max}}{B_{\min}} = (\sqrt{2})^{N-1}$$

where B_{\max} is the brightest spot on the display, B_{\min} is the darkest spot on the display and N is the number of gray shades.

The display brightness control was adjusted downward until the brightness of the active display dark area was just masked by ambient light reflection (B_{\min}). This area was then measured with a spot photometer. The brightest spot on the display (B_{\max}) was then measured at the largest target image and the display controls were adjusted to produce the required B_{\max}/B_{\min} ratio.* This allowed the full video signal to be displayed over the specified contrast range. No compensation was made for the normalized aperture response of the display as the target was reduced in size from 20 to 4 TV lines.

* This method for determining and adjusting contrast is more appropriate than Johnson's technique of eliminating steps on a step wedge displayed with the test image. Eliminating steps on a step wedge is really contrast or dynamic range clipping and not range compression or contrast reduction. As a result, information available at the clipped end of the step wedge is lost beyond recovery. If, however, gray shade reduction is accomplished by scale compression, the information is still present in the display, although at a reduced brightness contrast. This compression effect is analogous to the contrast washout that occurs in a cockpit display. In both cases the utility of the display is dependent on the observer's ability to perceive and use the low brightness contrast information.

Ambient lighting was held at 130 foot candles on the table surface and 90 foot candles incident on the CRT face throughout the experiment. This level of illumination is equivalent to sunlight through medium overcast.

PROCEDURE

The observer was seated 26 inches from the TV display and allowed to read the instructions and study photo keys of the target aircraft. These keys consisted of 3 large-scale photos of the 4 model aircraft, representing three different views of the aircraft, front, top and front oblique. The test procedure consisted of a 15-second presentation of the target image on the display, followed by a 10-second pause and then the next image. This process continued until a full set of 20 images, 4 aircraft at 5 TV lines-over-target values were presented. To insure a basic familiarity with the procedure each observer was presented this set of images at 7 shades of gray as practice.

After viewing a full set of images at one gray shade level (20 images), the TV system was calibrated for another gray shade level and the observer was re-tested. This process was repeated until all three gray shade levels (3, 5, and 7) were presented. The order of presentation of gray shades and images was counterbalanced across observers.

The timing operations were controlled by an automated clock-relay system that operated a mechanical shutter. The observer was instructed to respond to all images even if he was only guessing. He could respond at any time during the

15-second display period by depressing a response switch and calling out the aircraft number on the photo key. Depressing the response switch closed the shutter and stopped the clock. The response accuracy and response time were then recorded by the experimenter, and the slide projector was advanced to the next slide.

Section 4

RESULTS

The number of correct responses and percent correct for each gray shade and TV line combination are presented in Figures 3 and 4. The summary table of the analysis of variance performed on the correct response data is shown in Table 1. This analysis indicated that significant differences were present in both gray shades and TV lines variables.

As can be seen in Figure 3, performance increases from 3 to 5 shades of gray and remains essentially the same from 5 to 7 shades of gray. A Newman-Keules⁶ test of the differences between shades of gray showed 3 gray shades to yield significantly poorer performance than 5 and 7 gray shades.

Performance increases with increases in TV lines over target until 12 lines over target and then levels out with further increases in TV lines (Figure 4). A Newman-Keules test of the differences between values for the TV line variable showed significant increases in performance from 4 to 8 to 12 lines and no differences between 12, 16 and 20 lines. The interactions between gray shades and TV lines was not significant (Figure 5).

Average times per response for gray shades and TV lines-over-target are shown in Figures 6 and 7. An analysis of variance (Table 2) showed differences at the .01 level in time per response for the TV lines variable and no differences for the gray shades variable.

The lack of significance for the gray shades variable on the time measure can be seen in Figure 6. The total difference in mean response time averaged over 120 responses is less than one second between 3 and 7 shades of gray.

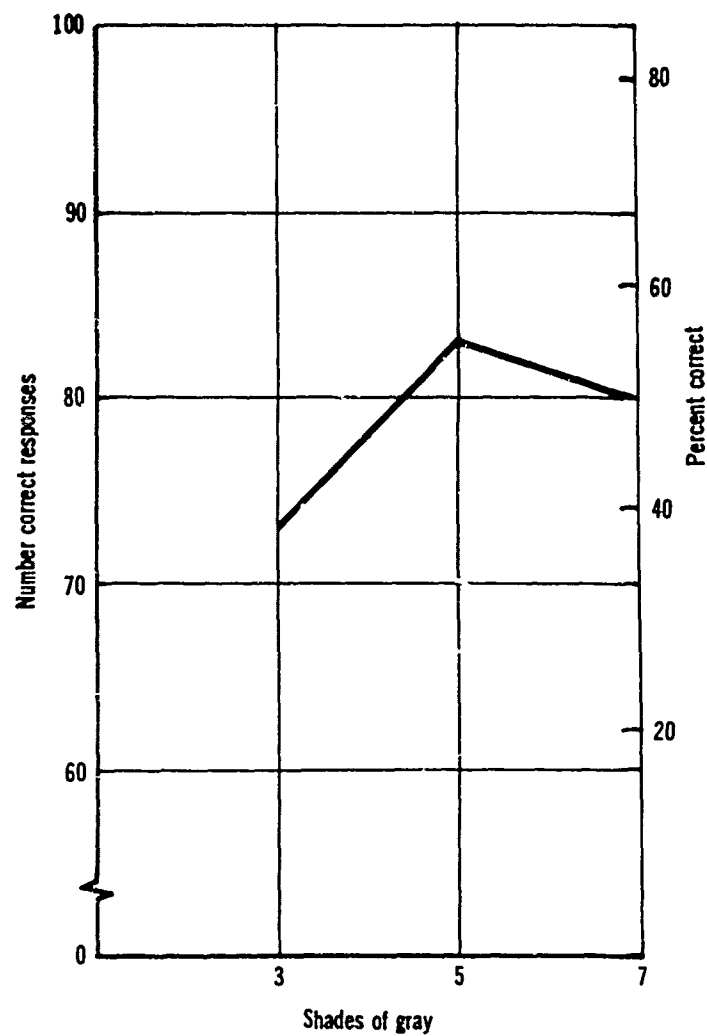


Figure 3 Percent and Number Correct for 3, 5 and 7 Shades of Gray

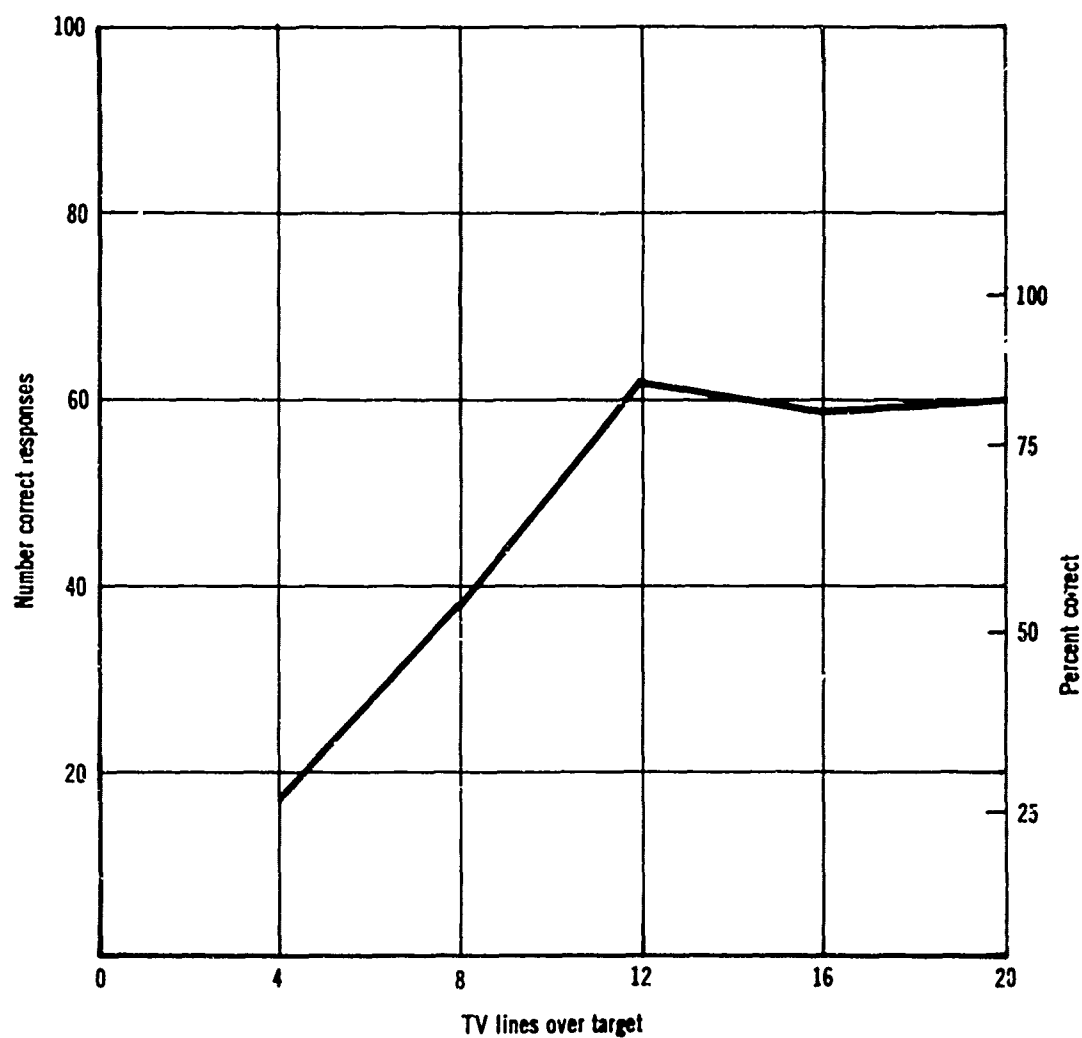


Figure 4 Percent and Number Correct Responses for TV Lines over Target

Table 1 Analysis of Variance (Number of Correct Responses)

Summary Table

Source of Variance	Degrees of Freedom	Mean Square	F
Gray Shades (G)	2	.88	5.17*
TV Lines (T)	4	21.09	20.88**
Subjects (S)	5	3.46	
GXT	8	.29	
GXS	10	.17	
TXS	20	1.01	
GXTXS	40	.51	

* Significant at the .05 level.

** Significant at the .01 level.

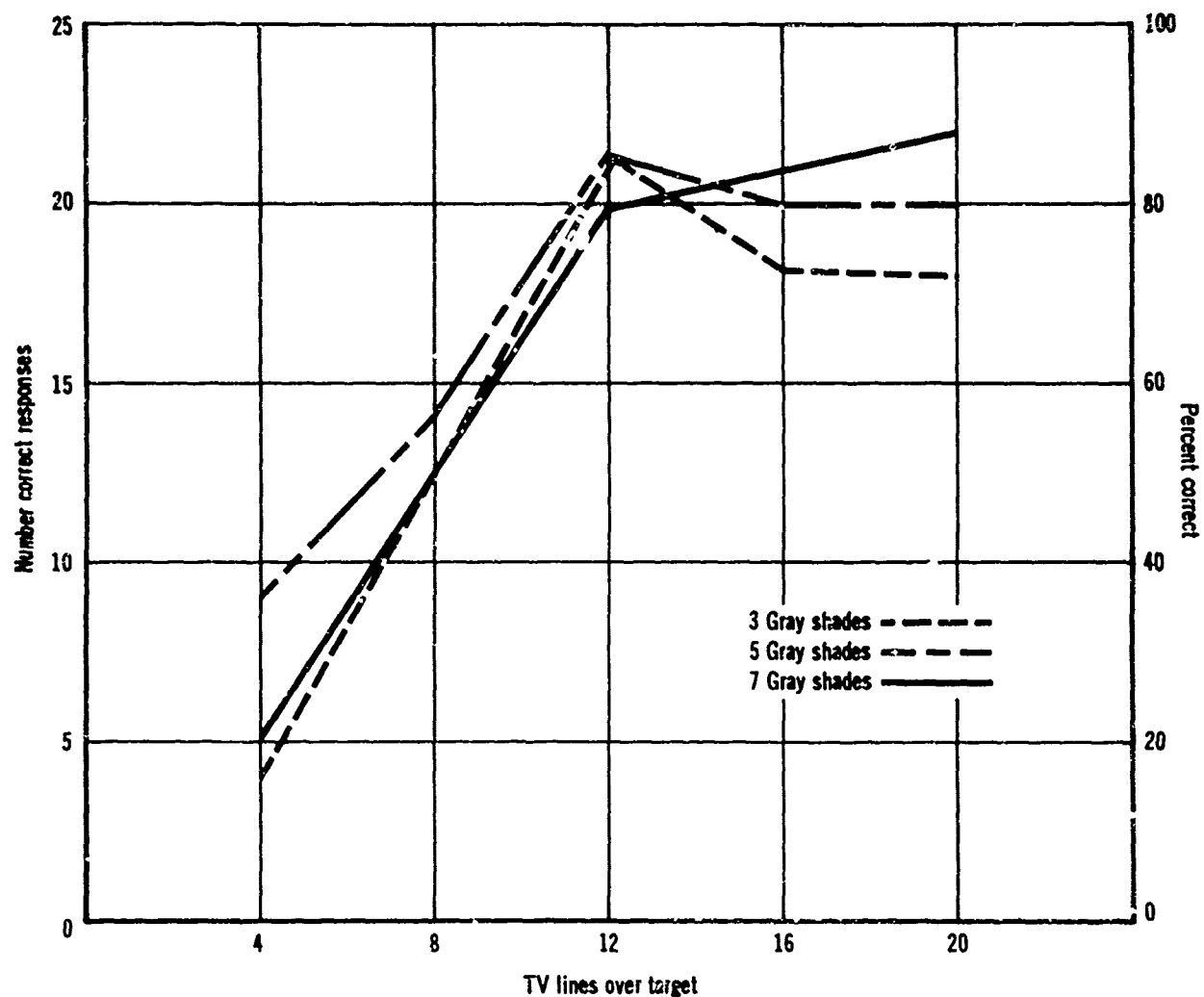


Figure 5 Percent and Number Correct Responses for TV Lines over Target
by Gray Shades Interaction

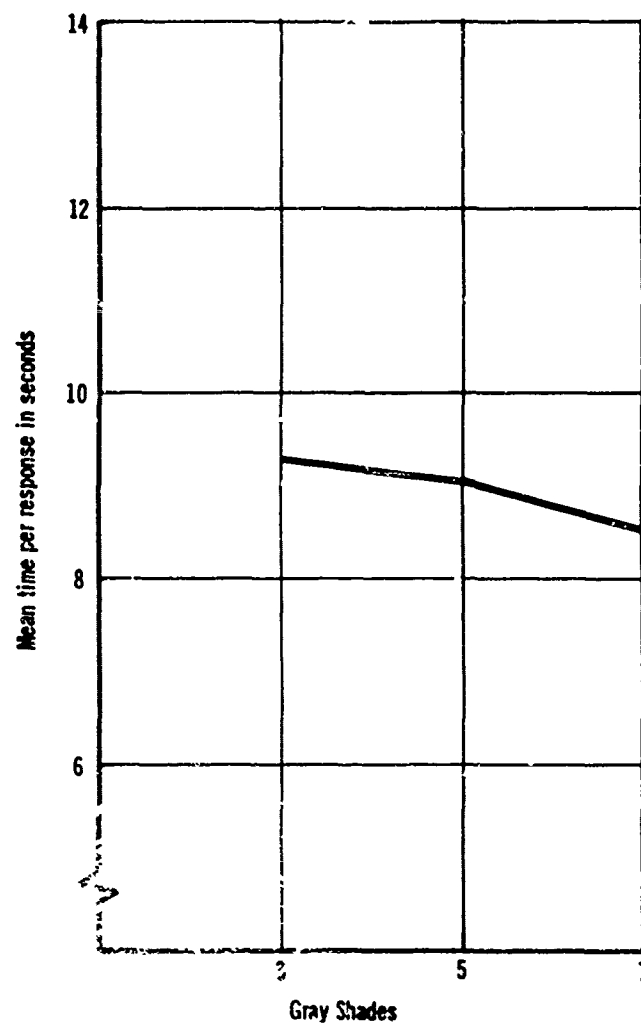


Figure 6 Mean Time per Response -- Gray Shades

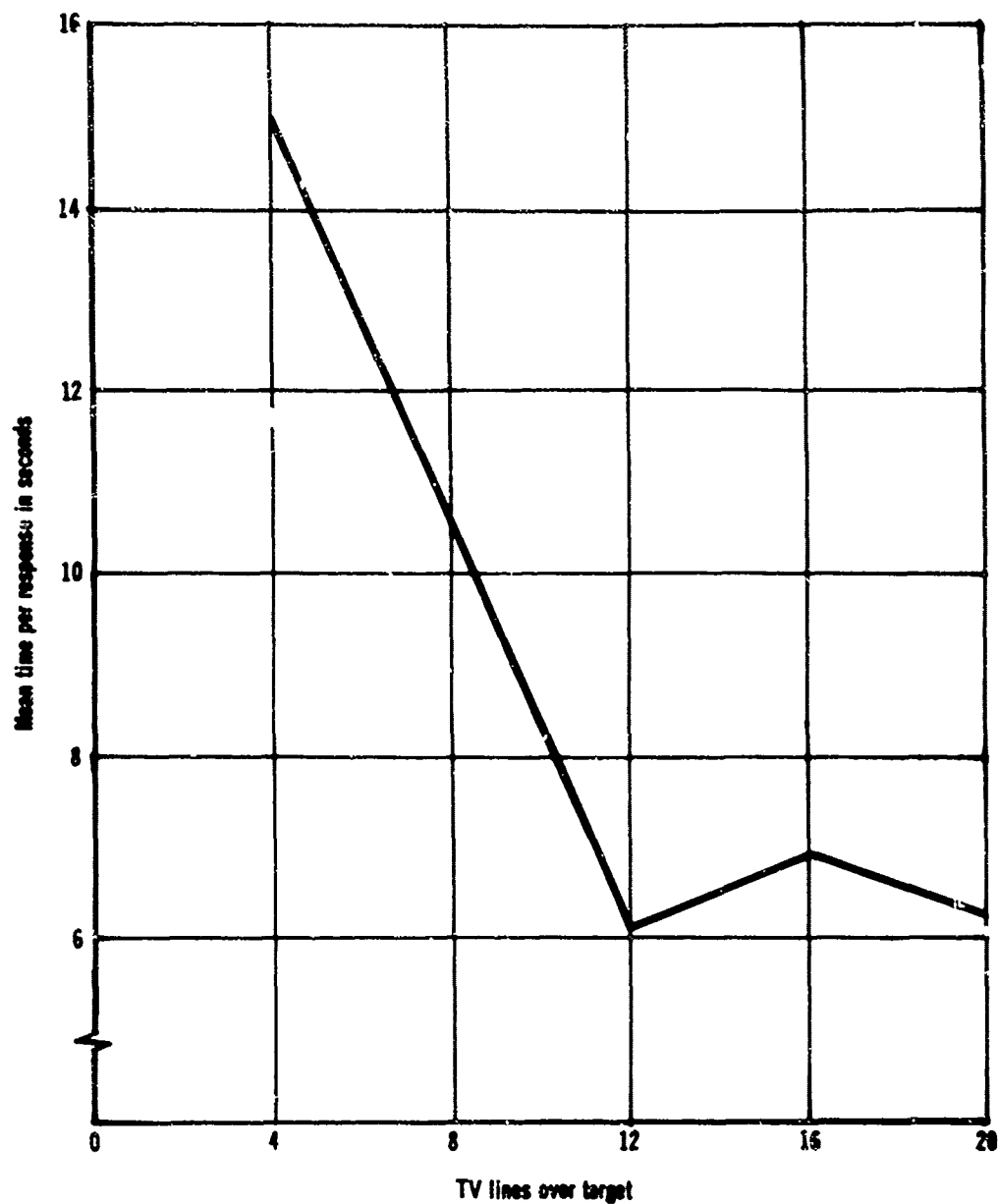


Figure 7 Mean Time Per Response - Lines Over Target

Table 2 Analysis of Variance (Response Time)

Summary Table

Source of Variance	Degrees of Freedom	Mean Square	F
Gray Scales (G)	2	74.44	1.39
TV Lines (T)	4	5034.41	35.55**
Subjects (S)	5	875.00	
GXT	8	25.05	
GXS	10	53.64	
TXS	20	141.61	
GXTXS	40	21	

** Significant at the .01 level.

Figure 7 shows the mean time per response decreasing with increases in lines over target up to 12 lines and then leveling off to a value of about 6.5 seconds per response. A Newman-Keules test for differences between TV line values showed the same pattern as was found for number correct, performance increased (response time went down) from 4 to 8 to 12 lines and showed no change from 12 to 16 to 20 lines. The interactions between gray shades and TV lines was, again, not significant (Figure 8).

In summary, the results of this experiment show that the number of correct responses decrease as you go from 5 and 7 to 3 shades of gray. The response time measure was not affected by gray shades variation. Varying TV lines-over-target yielded an overall performance increase (number correct goes up and response time goes down) as lines were increased from 4 to 12. Further increases in lines-over-target had no effect. There was no interaction between gray shades and lines-over-target for either response measure.

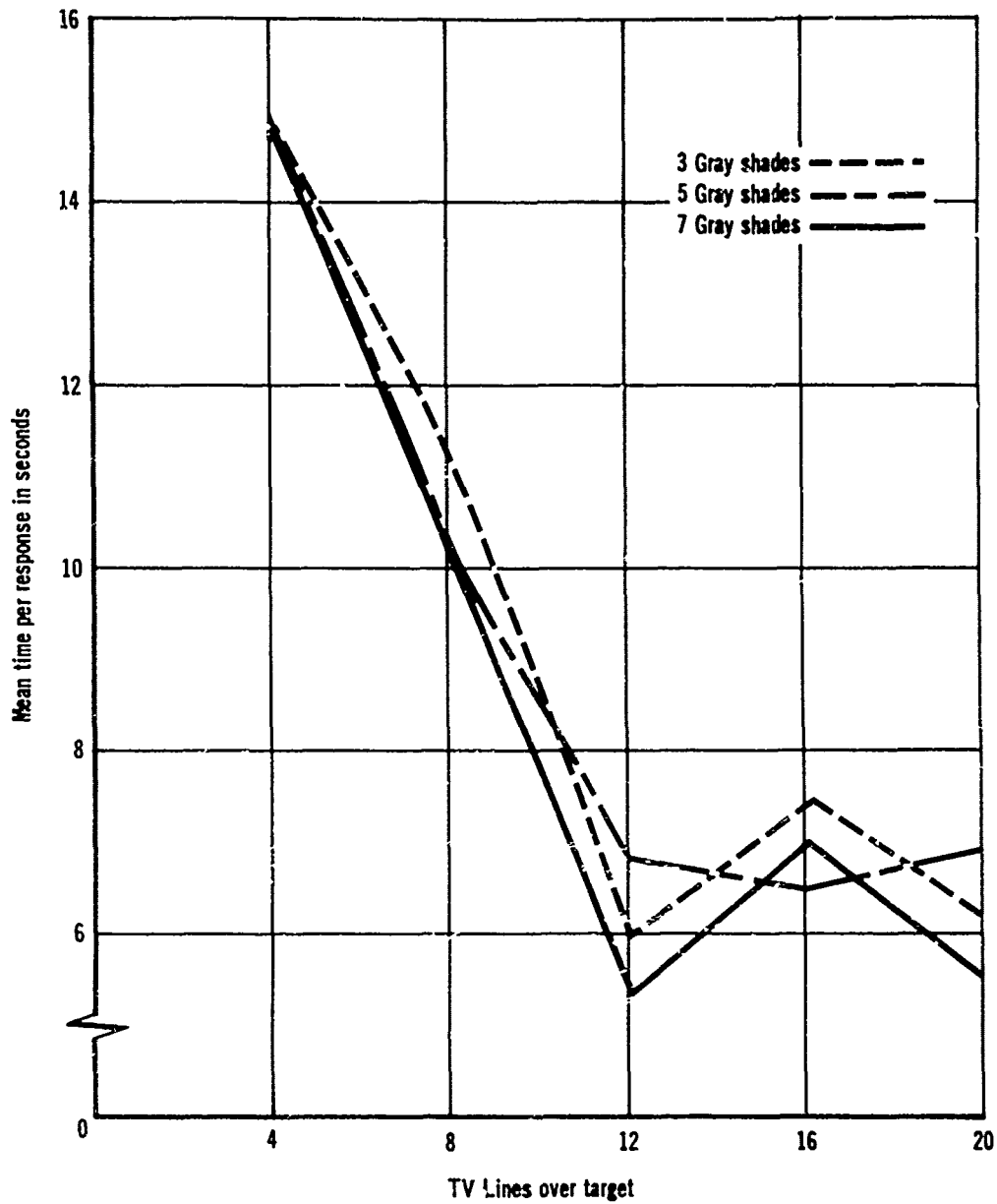


Figure 8 Mean Time per Response for TV Lines Over Target by Gray Shades Interaction

Section 5

DISCUSSION

The results of this study indicate that a minimum of 5 gray shades are required for adequate target identification on a cockpit TV display. The other studies in the area,^{1,2} although they use target material not as directly applicable to this problem, support this conclusion. This, in conjunction with the use of trained observers and a realistic target identification task, gives a high degree of face validity to the recommendation. The gray shade performance found here should generalize to similar displays requiring the same degree of gray discrimination. Generalization of the gray shade requirements can be made to IR displays of the type used in the Compass Eagle program and SLR displays as proposed for the Compass SLR and JIFDATS programs provided that the observers task is similar to that used in this study.

Five gray shades cover a contrast range of 300%, as calculated by the formula $\Delta B/B_{\min} \times 100$ where ΔB = difference in brightness between brightest and darkest area ($B_{\max} - B_{\min}$) and B_{\min} = the darkest area. A contrast difference as large as this contains many more discriminable shades of gray than would be indicated by the five $\sqrt{2}$ gray shades used to define the contrast range in the study.*

* About 14 discriminable gray tones were observed by Williston and Ellison (1949)⁷ for pulsed signals in noise background.

The data indicate that it would be profitable to include some measure of subjective gray differences such as Gray JNDs, along with the physical description of the contrast range given by the number of gray shades. In addition, data should be generated as to the number of discriminable grays required to identify a target. Such information would be of great use to anyone attempting digital processing of the TV display output, as it would give a good indication of the quantization levels needed to adequately define the gray rendition in the image.

The data on the number of TV lines-over-target agree fairly well with other experiments with respect to the shape of the performance curve for TV lines-over-target; however, performance for all TV lines over target was generally lower than that found in most other studies. To date, other experiments have used simple shapes to obtain their data.^{1,2,3,5} This would tend to increase performance at all TV lines-over-target, the simpler the target, the easier the task.

The Elite Look Program⁴ with its use of simulated tactical targets dubbed onto real imagery has come closest to simulating the type of imagery found on a real world display. In our study we attempted to even further simulate the display situation by using trained observers and an identification task using targets of a complexity commensurate with that found in a real world display. Our data and that found in the Elite Look Program show a good general agreement as to the minimum requirement for lines-over-target needed for effective performance. The best estimate of the number of lines-over-target for 70-80% correct identification would be about 12-16 lines-over-target. This data is

derived from our experimental work and the Elite Look Program, both of which used complex imagery for target material, and will apply to similar displays having a sensor system resolution greater than the display resolution and a target identification task of equal difficulty. This data indicates that TV line requirements for target identification increase with increasing target complexity. Because of this a great deal of caution should be used in generalizing from laboratory studies using target material greatly different from that expected in the real world displays.

To clarify specific data points and to generally increase the validity of the results, additional work needs to be done to expand the number of targets, subjects, and aircraft attitudes used in the experiment. Further work should also be done in the number of discriminable shades of gray required to identify an air-to-air target, closing rates between aircraft, and the effect of training on performance.

As these tests were performed at the highest S/N ratios commensurate with the TV equipment used, studies manipulating video signal variables, such as random noise and signal dynamic range, while maintaining constant display capability are also recommended. The results of such studies will determine gray shade interactions with signal variables of the type found in actual real-time, air-to-air displays.

REFERENCES

1. Erickson, R. A., Linton, P. M. and Hemmingway, J. C., Human Factors Experiments with Television. USN:NWC TP 4573, October 1968.
2. Johnson, Dorothy M., Target Recognition on TV as a Function of Horizontal Resolution and Shades of Gray. Human Factors, June 1968, 201-210.
3. Baker, C. A. & Nicholson, R. M., Raster Scan Parameters and Target Identification. Proceedings of the 19th Annual Aerospace Conference (NAECON) 1967, P285-290.
4. Improved Reconnaissance Display Techniques (Elite Look) Vol 1 Technical Studies and Results. Air Force Avionics Laboratory, Wright-Patterson AFB, Ohio (AFAL-TR-68-105), 1968.
5. Hemmingway, J. C. and Erickson, R. A., Relative Effects of Raster Scan Lines and Image Subtense on Symbol Legibility on Television. Human Factors, August 1969, 331-338.
6. Winer, B. J., Statistical Principles in Experimental Psychology, McGraw Hill, New York, 1962, pp 77-89.
7. Williston, S. S. and Ellison, J. V. (1949), Personal communication to authors.